

# **METHOD AND APPARATUS FOR MPEG-4 FGS PERFORMANCE ENHANCEMENT**

## **FIELD OF THE INVENTION**

[0001] The present invention generally relates to fine granularity scalable codec, and  
5 more specifically to the architecture, prediction mode and bit allocation of fine  
granularity scalable codec.

## **BACKGROUND OF THE INVENTION**

[0002] Applications of multimedia are more and more popular in today's world. For  
instance, one can listen to a CD player or access a web page via the Internet. One of the  
10 common problems in multimedia applications via the Internet is that the data of  
uncompressed video is too large for storage and transmission. Several coding standards  
have been defined by ITU-T and ISO-IEC MPEG committees to address data  
compression issues. With the establishment of these standards, it is much easier to store  
and transmit video data.

15 [0003] Because the Internet technology has advanced greatly over the past few years,  
one can read a web page, play games, and download files over the Internet nowadays.  
Streaming video is an important web application. People can access pre-encoded video  
clips from a video server via the network. The greatest advantage of streaming video is  
people can subscribe the video data through the Internet connection from anywhere. In  
20 streaming video, users may access videos from heterogeneous networks such as ADSL,  
cable modem, etc. Due to the bandwidth variations, the streaming video provider must  
transmit the bitstream at variable bit-rates.

[0004] There are some traditional methods for bit-rate adaptation. One is to encode multiple bitstreams at the encoding time. However, in video multicast environment, hundreds or thousands of clients may access the data at the same time. The total bit rate required is the sum of the bit rates of these multiple bitstreams. Another is to encode the bitstream at a highest bit-rate of the Internet and then transcode the bitstream into different bit-rates. First, the transcoder decodes the encoded bitstream, and then re-encodes it to meet the bit-rate that is suitable for each client. In this way, the streaming video provider can use a transcoder to transcode the bitstream into different bit-rates for different users.

[0005] A new concept called Fine Granularity Scalability (FGS) was proposed and standardized in MPEG-4 Draft Amendment 4. FGS contains one base layer and one enhancement layer. The FGS base layer is generated using an MPEG-4 coder at the lowest bit rate of all possible connections. FGS takes the original and reconstructed discrete cosine transform (DCT) coefficients to generate the enhancement layer bitstream using bit-plane coding. The reconstructed DCT coefficients are subtracted from the original ones to generate the residues introduced by the quantization process. Then the FGS codec uses bit-plane coding to encode these residues and outputs these bit-planes from the most significant bit (MSB) to the least significant bit (LSB). The enhancement layer can be truncated at any amount of bits. If the client has extra bandwidth after receiving the FGS base layer, it can also receive the enhancement layer. The more the FGS enhancement bit-planes are received, the better the reconstructed quality is. FGS provides a bit-rate range from the base-layer bit-rate to the upper bound of the client bandwidth. Therefore FGS is very suitable for streaming video with multicasting. As

shown in Fig. 1, all clients (client 1, 2, 3) can receive the FGS base layer at minimum perceptual quality. Because of insufficient bandwidth, client 1 can not receive the FGS enhancement layer. But client 2 and client 3 can receive the FGS bit-planes as many as they can.

5   **[0006]**   Because FGS can support a wide range of bit-rates to adapt to bandwidth variations, it is much more flexible than other coding schemes for streaming video applications. Therefore FGS becomes more and more popular in streaming video applications. While providing such a high flexibility for bandwidth adaptation, the coding efficiency of an FGS coder is not as good as that of a non-scalable coder at the same bit-  
10   rate. The inefficient coding performance mainly results from two factors. First, only coarse predictions are used for the motion-compensated predictive coding of the FGS base-layer, while the coding residuals (the image details) reconstructed from the enhancement-layer are not used for prediction. Second, there is no motion-compensated prediction loop involved in the FGS enhancement-layer coder. That is, each FGS  
15   enhancement-layer frame is intra-layer coded. Since the FGS base-layer is encoded at the lowest bit-rate with the minimal human perceptual visual quality, the coding gain in the temporal prediction of the FGS base layer is usually not as good as that for a non-scalable coder.

**[0007]**   Fig. 2 shows the encoding process to produce the FGS base-layer and  
20   enhancement-layer bitstreams. The base layer is encoded using an MPEG-4 non-scalable coder at bit-rate  $R_b$ . The FGS enhancement-layer coder uses the original and the de-quantized DCT coefficients as its inputs and generates the FGS enhancement-layer bitstream using bit-plane coding. The encoding procedure of the FGS enhancement-layer

bitstream goes as follows. First, the de-quantized DCT coefficients are subtracted from the original DCT coefficients to obtain the quantization residues. After generating all DCT residues of a frame, the enhancement-layer coder finds the maximum absolute value of these DCT residues to determine the maximum number of bit-planes for this frame.

- 5 After defining the maximum number of bit-planes in a frame, the FGS enhancement-layer coder will output the enhancement data bit-plane by bit-plane started from the most significant bit-plane (MSB plane) to the least significant bit-plane (LSB plane). The binary bits in each bit-plane are converted into symbols, and variable length encoded to generate the output bitstream. The following example illustrates the procedure, where the  
10 absolute quantization residues of a DCT block are given as follows:

5, 0, 4, 1, 2, 0, ... 0, 0

[0008] The maximum value in this block is 5 and the number of bits to represent 5 in a binary format (101) is 3. Writing every value in binary format, the 3 bit-planes are formed:

15                   1, 0, 1, 0, 0, 0 ... 0, 0       (MSB)

                  0, 0, 0, 0, 1, 0 ... 0, 0       (MSB-1)

                  1, 0, 0, 1, 0, 0 ... 0, 0       (LSB)

- [0009] Fig. 3 illustrates the FGS decoding process for the enhancement-layer frame reconstruction. The process of decoding the FGS base layer is the same as that of  
20 decoding an MPEG-4 non-scalable bitstream. Due to the embedded characteristics of FGS streams, the decoder receives and variable-length decodes the bit-planes of DCT residues from the MSB bit-plane to the LSB bit-plane. Because the decoder may not

receive all blocks of some specific bit-plane, the decoder fills 0's into the non-received blocks of bit-planes and performs IDCT to convert the received DCT coefficients into the pixel values. These pixel values are subsequently added to the base-layer decoded frame to obtain the final enhanced video image.

5    **[0010]**    Although FGS can support a wide range of bit-rates to ease the adaptation of channel variations, it, however, presents some disadvantages. Referring to Fig. 2, the input signal fed into the enhancement-layer coder is the quantization error of the prediction residue of the incoming video with reference to its base-layer reconstructed version, which is encoded at the lowest bit-rate with the minimum visual quality. In this way, the base-layer video is usually not able to approximate the incoming video with  
10    high accuracy, so the quantization error is relatively large, thereby leading to low coding efficiency. The performance of single-layer coding is better than the FGS coding at the same transmission bit-rate because the single-layer coding uses the full-quality video for prediction. The performance degradation can be up to 1.5 to 2.5 dB as reported in the  
15    prior arts.

**[0011]**    To overcome this problem, there have been several relevant works proposed for enhancing the visual quality of FGS coding as will be briefly described below.

**[0012]**    A method to improving the FGS coding efficiency, referred to as "Adaptive Motion Compensated FGS" (AMC-FGS) has been proposed. The AMC-FGS codec is  
20    featured with two simplified scalable codecs: one-loop and two-loop MC-FGS with different degrees of coding efficiency and error resilience. The two-loop MC-FGS employs an additional MCP loop at the enhancement-layer coder for only B-frames to obtain better coding efficiency. Since B-frames are not referenced by other frames for

prediction during encoding and decoding, there will be no error propagation due to the loss of B-frame data. If drifting errors occur in one B-frame, the drifting errors will not propagate to the following frames. The one-loop MC-FGS introduces fine predictions for P- and B-frames, leading to relatively higher coding efficiency compared to the two-loop MC-FGS. However, the error robustness would become significantly lower since the drifting error can be rather significant if the enhancement-layer data used for prediction of the base layer of P-frames cannot be received at the decoder due to packet losses caused by insufficient channel bandwidth or channel error, leading to significant quality degradation. An adaptive decision algorithm is used in AMC-FGS to dynamically switch over the two prediction schemes to achieve better tradeoff in terms of coding efficiency and error robustness.

**[0013]** A new FGS structure which is called “Progressive FGS (PFGS)” has also been proposed. In the proposed structure, the enhancement layer not only can refer to the FGS base layer but also can refer to the previous enhancement-layer data. However, the same drifting errors also confuse the output quality if referenced bit-planes can not be guaranteed to transmit to the decoder when the bandwidth is dropped.

**[0014]** Another method that has been proposed is referred to as “Robust Fine Granularity Scalability (RFGS)”. The method focuses on the tradeoff between coding efficiency and robustness by adopting additional motion compensation (MC) loop at the enhancement layer and including leaking prediction. The extra MC loop can improve the coding efficiency by referencing high quality frame memory, and the accompanied drift errors are handled by leaking prediction. A leaky factor  $\alpha$  ( $0 \leq \alpha \leq 1$ ), which is bound with the estimated drift errors, is introduced into the reconstructed frame memory at the

enhancement layer. And, a separated factor introduced is the number of referenced bit-planes  $\beta$  ( $0 \leq \beta \leq$  maximal number of bit-planes) which is utilized in partial prediction. By adjusting both factors, the RFGS can provide flexibility of various encoding schemes. If the leaky factor ( $\alpha$ ) is set to zero, it is almost the same as the original FGS. If the factor  
5 ( $\alpha$ ) is set to unity for all referencing frames, the prediction modes of RFGS and MC-FGS are equal.

### **SUMMARY OF THE INVENTION**

[0015] This invention has been made to enhance the performance of the fine granularity scalable codec. The primary object of this invention is to provide a new  
10 architecture of FGS codec with three prediction modes that can be adaptively selected. Another object of the invention is to provide a method to adaptively select a prediction mode for each macroblock of input signals. It is yet another object to provide a method of enhancement-layer bit-plane truncation for the FGS codec.

[0016] According to the invention, both the encoder and the decoder of the fine  
15 granularity scalable codec have a base layer which comprises a coarse prediction loop with a base layer mode selector, and an enhancement layer which comprises a fine prediction loop with an enhancement-layer mode selector. The base-layer mode selector can be controlled to select the output of either coarse or fine prediction for the base layer. Similarly, the enhancement-layer mode selector can also be controlled to select the output  
20 of either coarse or fine prediction for the enhancement layer.

[0017] Three prediction modes are provided for the fine granularity scalable codec of this invention. The codec operates in an all-fine prediction mode when both the base-

layer mode selector and the enhancement-layer mode selector are switched to select the fine prediction output, in an all-coarse prediction mode when both the base-layer mode selector and the enhancement-layer mode selector are switched to select the coarse prediction output, and in a mix prediction mode when the base-layer mode selector is switched to select the coarse prediction output and the enhancement-layer mode selector is switched to select the fine prediction output.

[0018] The prediction modes of the encoder are adaptively selected for each macroblock of the input video signals. A two-pass encoding procedure is adopted in this invention. In the first-pass encoding, the encoding parameters of all macroblocks are collected, including prediction error values of fine and coarse predictions, and best-case and worst-case estimated mismatch errors introduced with the fine prediction in the case that the enhancement layer data used for prediction cannot be received at the decoder. A coding gain is derived from the fine and coarse prediction error values and a predicted mismatch error is derived from the best-case and worst-case estimated mismatch errors.

15 A coding efficiency metric defined as the ratio of the coding gain over the predicted mismatch error is computed for each macroblock. The mean and standard deviation of the coding efficiencies from all the macroblocks in a frame are also computed.

[0019] The macroblocks are then classified into three groups based on the coding efficiency of each macroblock. The macroblocks of each group are assigned and encoded with an identical prediction mode. A macroblock is encoded with the all-coarse prediction mode if the coding efficiency of the macroblock is smaller than the difference of the coding efficiency mean and a pre-determined multiple of the coding efficiency standard deviation, and the macroblock is encoded with the all-fine prediction mode if the



coding efficiency of the macroblock is larger than the sum of the coding efficiency mean and the pre-determined multiple of the coding efficiency standard deviation. Otherwise the macroblock is encoded with the mix prediction mode.

5 [0020] A new rate adaptation algorithm is further provided for truncating the enhancement-layer bit-planes with three different cases of available bandwidths: low bit-rate, medium bit-rate and high bit-rate. In the low bit-rate case, the enhancement-layer bit-planes of I/P-frames are truncated as much as possible. The bit allocation is made only for I/P-frames while the enhancement layer data of B-frames are all dropped in truncation. In the medium bit-rate case, excessive bits are distributed to B-frames after  
10 the bit allocations to I/P-frames can guarantee the bit-planes of I/P-frames used for fine prediction can be completely sent. In the high bit-rate case, the number of bits for distribution is controlled by the size of bit-planes and varies at particular bit-rates. To avoid a large variation between two neighboring frames if no more bits are allocated to I/P-frames, the distributed bit-allocations among frames should be balanced.

15 [0021] The foregoing and other objects, features, aspects and advantages of the present invention will become better understood from a careful reading of a detailed description provided herein below with appropriate reference to the accompanying drawings.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

20 [0022] Fig. 1 shows how FGS bitstreams are transmitted to different clients with different bandwidth.

[0023] Fig. 2 shows the encoding process to produce the FGS base-layer and

enhancement-layer bitstreams.

[0024] Fig. 3 shows the decoding process for the FGS base-layer and enhancement-layer frame reconstruction.

[0025] Fig. 4 shows the encoder structure of the novel FGS codec with inter-layer  
5 prediction according to the present invention.

[0026] Fig. 5 shows the decoder structure of the novel FGS codec with inter-layer prediction according to the present invention.

[0027] Fig. 6 shows the encoder structure of the novel FGS codec with inter-layer prediction in which the base layer only has coarse prediction according to the present  
10 invention.

[0028] Fig. 7 shows the decoder structure of the novel FGS codec with inter-layer prediction in which the base layer only has coarse prediction according to the present invention.

[0029] Fig. 8 shows the two-pass encoding procedure of this invention.

15 [0030] Fig. 9 shows an example distribution and the relationship between the estimated mismatch errors and coding gains for a number of MBs.

[0031] Fig. 10 shows the performance comparison of the method of this invention to three other conventional methods using Mobile test sequence.

[0032] Fig. 11 shows the performance comparison of the method of this invention to  
20 three other conventional methods using Coastguard test sequence.

[0033] Fig. 12 shows the frame-by-frame performance comparison of the method of

this invention to three other conventional methods using Coastguard test sequence with a base-layer bit-rate of 384 kbps and an enhancement layer bit-rate of (a) 0 kbps, (b) 256 kbps and (c) 768 kbps respectively.

[0034] Fig. 13 shows the frame-by-frame performance comparison of the method of this invention to three other conventional methods using Mobile test sequence with a base-layer bit-rate of 512 kbps and an enhancement layer bit-rate of (a) 0 kbps, (b) 256 kbps and (c) 768 kbps respectively.

[0035] Fig. 14 shows the 4<sup>th</sup> decoded picture with 512 kbps at base layer and 512 kbps at enhancement layer by (a) the original FGS encoder (27.5 dB) and (b) the Hybrid MB-MSFGS method of this invention (32.4 dB).

#### **DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS**

[0036] Figs. 4 and 5 depict the block diagrams of the novel three-mode FGS codec according to the present invention. As shown in Fig. 4, the encoder structure comprises an enhancement layer and a base layer. The enhancement layer has a DCT unit 401, a bit-plane shift unit 402, a maximum value finder 403, a bit-plane variable length coder 404, and a fine prediction loop which includes a bit-plane divider 405, an IDCT unit 406, a fine frame memory 407 and a motion compensation unit 408 with a switch SW1 for configuring the prediction modes in the enhancement layer. The base layer has a DCT unit 411, a quantization unit 412, a variable length coder 413 and a coarse prediction loop which includes an inverse quantization unit 414, an IDCT unit 415, a coarse frame memory 416, a motion estimation unit 417, a motion compensation unit 418 with a switch SW2 for configuring the prediction modes.

[0037] The decoder structure of this invention as shown in Fig. 5 also comprises an enhancement layer and a base layer. The enhancement layer has a bit-plane variable length decoder 501, a first IDCT unit 502, and a fine prediction loop which includes a bit-plane divider 503, a second IDCT unit 504, a fine frame memory 505 and a motion compensation unit 506 with a switch SW3 for configuring the prediction modes in the enhancement layer. The base layer has a variable length decoder 510, an inverse quantization unit 511, a third IDCT unit 512, and a coarse prediction loop which includes a coarse frame memory 513, and a motion compensation unit 514 with a switch SW4 for configuring the prediction modes in the base layer.

[0038] The principle and operation of the basic fine granularity scalable codec used in this invention have been well known and described in the prior art. The architecture of the novel FGS codec of this invention provides switches SW1, SW2, SW3 and SW4 for adaptively selecting three prediction modes to improve coding efficiency and performance. The following will describe the principles of various prediction modes and their operations.

[0039] As shown in Fig. 4, the encoder contains two switches, SW1 and SW2, for configuring the prediction modes of the two motion-compensated prediction loops in the enhancement-layer (EL) and base-layer (BL) coders, respectively. The upper switch SW1 is used to configure the prediction from either of fine and coarse memories for the motion-compensation loop at the EL coder; while SW2 is for choosing the BL's prediction mode (SW = 1: fine prediction; SW = 0: coarse prediction). As summarized in Table 1, three coding modes are provided in the encoder at the macroblock (MB)-level according to this invention: All-Fine Prediction (AFP: SW1 = 1 and SW2 = 1), All-

Coarse Prediction (ACP: SW1 = 0 and SW2 = 0), and Mix Prediction (MP: SW1 = 1 and SW2 = 0).

[0040] According to this invention, the prediction modes of the encoder are adaptively selected for each macroblock of the input video signals by the mode selection switches SW1 and SW2 that are controlled by a mismatch estimation and mode decision unit 419 as illustrated in Fig. 4. Both best-case and worst-case estimates of mismatch errors are computed in the mismatch estimation and mode decision unit 419 for making mode decision. Therefore, in addition to the best-case coarse prediction output from the motion compensation unit 418, a worst-case coarse prediction output  $PX_{BL2}^i$  is also provided by a worst-case base-line decoder 420. The method for adaptively selecting the prediction modes will be described in detail later.

[0041] One or two variable length coded (VLC) bits per MB are sent to the decoder to signal the prediction mode used. These coding modes have different characteristics in terms of coding efficiency and error robustness. If the AFP mode is selected, both BL and EL exploit predictions from the fine frame memory, leading to the highest coding efficiency. This, however, runs a high risk of introducing drifting error because the receivers may not be able to completely receive the EL bit-planes used in the fine predictions due to insufficient channel bandwidth or packet losses. As a whole, the operations in this mode are very similar to the one-loop motion-compensated FGS (MC-FGS). On the contrary, same as the baseline FGS, the ACP mode uses coarse predictions for both BL and EL. This mode guarantees no drifting error should the base-layer bitstream be received completely but its coding efficiency is the lowest among the three modes. The MP mode compromises on the coding efficiency and error robustness. It

adopts fine predictions for the EL and coarse predictions for the BL, respectively. With this mode, drifting error may occur at the EL when part of EL bit-planes used for fine predictions is lost; while the BL can be drift-free under the assumption that the decoder receives the whole BL data.

5   **[0042]**   In addition to the novel three-mode FGS codec, as a special case of the three-mode codec, another simplified FGS codec with only MP and ACP coding modes reduces the drift while sacrificing some coding gain introduced by the AFP coding mode. Without the AFP coding mode, the new codec reduces to the coder and decoder architectures shown in Figs. 6 and 7, respectively. This two-mode version is referred to as  
10   the “low-drift” mode, in contrast to the “high-gain” mode for the three-mode version. In this new codec, the overhead of sending the coding mode is reduced to one bit per MB.

Table 1 summarizes the prediction modes used in the codec of this invention.

Table 1. Three prediction modes used in the FGS coding scheme of this invention

| Prediction Modes                                | VLC Code                         | Description   |
|---|----------------------------------|---|
| All-Coarse Prediction<br>(SW1 = 0 and SW 2 = 0) | Low-drift: 1<br>High-gain: 10    | <ul style="list-style-type: none"> <li>Coarse prediction is used for both the base and enhancement layers. Same with original FGS.</li> <li>Strong error resiliency, but less coding efficiency</li> </ul>        |
| All-Fine Prediction<br>(SW1 = 1 and SW 2 = 1)   | Low-drift: N.A.<br>High-gain: 10 | <ul style="list-style-type: none"> <li>Fine prediction is used for both the base and enhancement layers. Same with one-loop MC-FGS.</li> <li>Highest coding efficiency, but sensitive to drift errors.</li> </ul> |
| Mix Prediction<br>(SW1 = 1 and SW 2 = 0)        | Low-drift: 0<br>High-gain: 0     | <ul style="list-style-type: none"> <li>Fine prediction is used for the enhancement-layer and coarse prediction for the base layer. Same with PFGS.</li> <li>Limit the drifting error at the base</li> </ul>       |

|  |  |  |
|--|--|--|
|  |  | layer, and achieve higher coding efficiency at high bit-rate than "Original FGS" |
|--|--|--|

**[0043]** According to this invention, to avoid performing motion re-estimation and sending one extra motion vector for each MB, the motion vectors obtained from the BL encoder are reused for the motion-compensation operation at the EL coder. However, the BL motion vectors may not be optimal for encoding the EL bitstream.

5 **[0044]** As discussed above, encoding with the coarse prediction (i.e., the ACP mode) is usually less efficient than that with the fine prediction (i.e., the AFP and MP modes), while drifting error may occur if the fine prediction is utilized but some of EL bit-planes used for prediction are not received by the decoder. This invention develops a statistical approach to estimating the best choice of prediction mode when the user bit-rates are  
10 unknown prior to the encoding.

**[0045]** As illustrated in Fig. 8, a two-pass encoding procedure is adopted in this invention. While performing the first-pass encoding, the encoding parameters of all MBs are collected, including the prediction error values with the fine and coarse predictions, respectively, and the estimated mismatch error introduced with the fine prediction in the  
15 case that the EL data used for prediction cannot be received at the decoder. Among these parameters, the difference between the prediction error values of the two predictions reflects their coding gain difference, while the mismatch error will result in error propagation to the subsequent frames. For example, the coding gain with the fine prediction can be significantly higher than that with the coarse one, which can be  
20 estimated as the difference between the fine and coarse prediction errors of the as follows:

$$G_i = \sum_{m=0}^{15} \sum_{n=0}^{15} \left( \|X_{\text{in}}^i(m, n) - PX_{\text{BL1}}^i(m, n)\| - \|X_{\text{in}}^i(m, n) - PX_{\text{EL}}^i(m, n)\| \right) \quad (1)$$

where  $X_{\text{in}}^i$  stands for the  $i$ th incoming MB;  $PX_{\text{BL}}^i$  and  $PX_{\text{EL}}^i$  represent the associated coarse and fine predictions of  $X_{\text{in}}^i$ , respectively. Note, the two norms in Eq. (1) represent the energy values (e.g., the magnitudes) of the two prediction errors with the fine and coarse prediction modes, respectively. A large  $G_i$  value for one MB implies that the fine prediction is much more accurate than the coarse one.

[0046] However, the coding gain comes with the risk of introducing drifting error because the fine prediction adopts part of EL data which may not be completely received at the decoder due to insufficient bandwidth or packet loss. In order to capture such drifting effect, the following two mismatch estimates are evaluated:

$$D_i^{\text{B}} = \sum_{m=0}^{15} \sum_{n=0}^{15} \|PX_{\text{BL1}}^i(m, n) - PX_{\text{EL}}^i(m, n)\| \quad (2)$$

$$D_i^{\text{W}} = \sum_{m=0}^{15} \sum_{n=0}^{15} \|PX_{\text{BL2}}^i(m, n) - PX_{\text{EL}}^i(m, n)\| \quad (3)$$

where  $D_i^{\text{B}}$  and  $D_i^{\text{W}}$  stand for the best-case and worst-case estimates of mismatch errors, respectively, under the assumption of zero motion-vector error concealment being used.  $PX_{\text{BL2}}^i$  is the coarse prediction from another BL coder which encoded at the base-layer bit-rate (i.e., without receiving any EL bits). The mismatch estimates indicate the bounds of concealment error. The best-case estimate  $D_i^{\text{B}}$  evaluates the lower bound of mismatch error since it assumes all the BL data in previous frames are received correctly. In contrast, the worst-case estimate  $D_i^{\text{W}}$  is to calculate the accumulated drift should the



decoder have only the base-layer (lowest) bandwidth. These two measures can be used to characterize the effect of drifting error, since they reflect the difference between the two frame memories of encoder and decoder. A MB with a large mismatch value implies that it is likely to result in more drifting error if lost.

- 5 **[0047]** Note that it is impossible to accurately estimate the actual mismatch while encoding without the knowledge about the channel bandwidths and conditions of client decoders. However, it is known that the actual mismatch error is bounded by these two estimates, that is,  $D_i^B \leq D_i \leq D_i^W$ . This invention uses the weighted average of these two estimates to predict the actual mismatch error:

$$10 \quad PD_i = k_D D_i^B + (1 - k_D) D_i^W \quad (4)$$

where  $k_D \in [0,1]$ . The selection of  $k_D$  is dependent on the distribution of decoder bandwidth.

- [0048]** In order to determine the coding mode of each MB so as to achieve good coding performance while keeping enough error robustness, a new index: “Coding gains  
15 Over Drifting Error” (CODE) is introduced:

$$CODE_i = G_i / PD_i \quad (5)$$

- where  $G_i$  and  $PD_i$  are obtained from Eqs. (1) and (4), respectively. The index in Eq. (5) can be used to characterize the relative gain of coding performance improvement over the potential drifting error for a MB coded with fine prediction. A large CODE value of a  
20 MB implies a high possibility that using the fine-prediction to encode the MB can achieve high coding gain while the potential drift penalty is not serious.

[0049] After extracting the features for all the MBs in one video frame, the mean and standard deviation of the “CODE” values,  $m_{\text{CODE}}$  and  $\sigma_{\text{CODE}}$ , are calculated as follows:

$$m_{\text{CODE}} = \frac{1}{N_{\text{MB}}} \sum_{i=1}^{N_{\text{MB}}} \text{CODE}_i \quad (6)$$

$$\sigma_{\text{CODE}} = \sqrt{\frac{1}{N_{\text{MB}}} \sum_{i=1}^{N_{\text{MB}}} (\text{CODE}_i - m_{\text{CODE}})^2} \quad (7)$$

5 where  $N_{\text{MB}}$  is the number of MBs in a frame.

[0050] The MBs are then classified into three groups which are encoded with distinct prediction modes (i.e., the ACP, AFP, and MP modes) using the two parameters as follows:

$$\text{MODE}_i = \begin{cases} \text{ACP} & \text{if } \text{CODE}_i < m_{\text{CODE}} - k\sigma_{\text{CODE}} \\ \text{AFP} & \text{if } \text{CODE}_i > m_{\text{CODE}} + k\sigma_{\text{CODE}} \\ \text{MP} & \text{otherwise} \end{cases} \quad (8)$$

10 [0051] Fig. 9 illustrates an example distribution of pairs of mismatch and coding gain for a number of MBs. The X-axis and Y-axis indicate the values of coding gain as defined in Eq. (1) and the mismatch error as in Eq. (4), respectively. The higher X-axis value states that the fine prediction is more beneficial for this MB by introducing more bits into the fine frame memory. In the case of adapting extra bits, the coding gain  
 15 accompanies the drifting error. Each spot on Fig. 9 stands for the (G,D) pair of one MB located in each category. The upper and lower solid straight lines represent (G,D) pairs with the CODE values of “ $m_{\text{CODE}} + k\sigma_{\text{CODE}}$ ” and “ $m_{\text{CODE}} - k\sigma_{\text{CODE}}$ ” ( $k = 1$  in this case), respectively; while the broken lines between them represent those with the value of

$m_{\text{CODE}}$ . Those MBs with  $(G,D)$  positions above the upper solid line are encoded with the AFP mode since this is expected to be likely to achieve significantly higher coding performance, while the drifting error introduced is not that serious if the decoder does not receive some of EL packets used for prediction. On the contrary, the MBs with  $(G,D)$  positions under the bottom solid line are encoded with the ACP mode since they are more sensitive to drifting error. The remaining MBs are encoded with the MP mode to achieve a better tradeoff between the coding gain and drifting error.

[0052] Because P-frames are used as the references for encoding the following B/P-frames, the prediction mode decision method of this invention is applied to P-frames. Moreover, B-frames will not be used as predictions for other frames, the drifting error will not propagate to other frames. Therefore the fine predictions are used aggressively to encode all MBs in B-frames.

[0053] While streaming, the streaming server truncates each EL frame to an appropriate size to fit the channel bandwidth of the client terminal. If the fine prediction is used for encoding the BL and EL, the bit-allocation scheme for truncating the FGS EL frames can influence the performance largely. For example, if reasonably more bits can be allocated to I/P-frames than B-frames, the decoder will be likely to receive more bit-planes of I/P-frames, leading to lower drifting error and higher video quality. In addition, B-frames can also reference to better-quality pictures for prediction at the encoder as well as for reconstruction at the decoder should more EL bit-planes of the reference pictures used for prediction be received.

[0054] In this invention, a new rate adaptation algorithm is presented for truncating the EL bit-planes at the video server with three different cases of available bandwidths:

low bit-rate, medium bit-rate, and high bit-rate. In the low bit-rate case, the available bandwidth is not sufficient to send all the EL bit-planes of I/P-frames used for the fine-predictions of both layers during the encoding process. Therefore, drifting error is inevitable when part of the EL data used for prediction is dropped in the truncation process. On the other hand, if the available bandwidth is high enough to send all the EL bit-planes used for fine-predictions, but is less than the bit-count of  $N_{BP}$  EL MSB bit-planes of all B-frame in a group of pictures (GOP), the excessive bits will be distributed among B-frames to balance the picture quality between I/P- and B-frames. Moreover, if the channel condition is even better, the surplus of bits will also be allocated among I/P-frames while the related bits are reserved to avoid drifting error. Such bit-rate adaptation by truncating the EL bit-planes can be performed at the server or routers. The truncation schemes for different cases are elaborated separately below. Table 2 describes the parameters used in the server bit-plane truncation algorithm of this invention.

Table 2. Parameters used for server rate adaptation

| Parameter                          | Description  |
|------------------------------------|--|
| $N_{GOP}$                          | the GOP size   |
| $N_{I\&P}$                         | the number of I- and P-frames in a GOP   |
| $N_B$                              | the number of B-frames in a GOP ( $N_B = N_{GOP} - N_{I\&P}$ )                 |
| <b>Pre-encoding at the encoder</b> |  |
| $N_{BP}$                           | number of bit-planes used for fine predictions while encoding                  |
| $PB_{EL}$                          | total number of EL bits in a GOP used for fine predictions                     |
| $PB_{I\&P,EL}$                     | number of EL bits in all I- and P-frames in a GOP used for the fine prediction |
| $PB_{B,EL}$                        | bit-count of $N_{BP}$ EL MSB bit-planes of all B-frame in a GOP                |
| $PB_{I\&P,EL}^n$                   | number of EL bits in the $n$ th I/P-frames in a GOP used for fine predictions  |

|   |  |
|---|--|
| $PB_{B,EL}^m$   | bit-count of $N_{BP}$ EL MSB bit-planes of the $m$ th B-frame in a GOP |
| <b>Parameters of bit-plane truncation at the server</b> |  |
| $TB_{EL}$   | bit-allocation of truncation for the EL in a GOP                       |
| $TB_{I\&P,EL}^n$  | bit-allocation of truncation for the $n$ th I/P-frames of EL in a GOP  |
| $TB_{B,EL}^m$   | bit-allocation of truncation for the $m$ th B-frame of EL in a GOP     |

### Case 1: low available bandwidth

[0055] In this case, the available channel bandwidth estimated at the server is less than the amount of EL bits of I- and P-frames used for the fine predictions while encoding. Since the available bandwidth is not sufficient to send all the bits used in fine prediction, this invention truncates the enhancement layers as much as possible for I- and P-frames. The truncation scheme for each I/P frame is adapted according to the number of bits used for prediction in each frame as follows:

$$TB_{I\&P,EL}^n = PB_{EL} \times \frac{PB_{I\&P,EL}^n}{\sum_{n=1}^{N_{I\&P}} PB_{I\&P,EL}^n} \quad (9)$$

10 [0056] In this case, the bit-allocation for B-frames are all set to be zero, that is,  $TB_{B,EL}^m = 0$ ,  $m = 1, 2, \dots, N_B$ . Eq. (9) is used if the current bit budget is less than  $PB_{I\&P,EL}$ . The bit-allocation is made only for I- and P-frames, while the EL data of B-frames are all dropped in truncation in this case. This strategy can achieve more robust performance at low bit-rates.

### 15 Case 2: medium available bandwidth

[0057] If the available bandwidth is sufficient for sending all the EL bits of I- and P-frames used for fine prediction, but is less than  $PB_{B,EL}$ , the server starts to distribute the excessive bits to B-frames after the bit-allocations to I/P-frames can guarantee the bit-planes of I/P-frames used for fine-prediction be completely sent to the receiver.

### 5 Case 3: high available bandwidth

[0058] If the available bandwidth is higher than that required for sending the number of EL bit-planes used for the fine prediction, the number of bits for distribution is controlled by the size of bit-planes and varies at particular bit-rates. However, when the bit-rate increases rapidly, there exists a large variation between two neighboring frames if  
 10 no more bits are allocated to I/P-frames. Therefore, the distributed bit-allocations among frames should be balanced to avoid large quality variations.

[0059] The EL bit-allocation algorithm according to this invention is summarized with a pseudo program below:

#### EL Bit-Allocation Algorithm

15 **Begin:**

**if** ( $TB_{EL} \leq PB_{I\&P,EL}$ ) /\* perform low-rate bit truncation \*/

$$TB_{I\&P,EL}^n = PB_{EL} \times \frac{PB_{I\&P,EL}^n}{\sum_{n=1}^{N_{I\&P}} PB_{I\&P,EL}^n}, \quad n = 1, 2, \dots, N_{I\&P};$$

$$TB_{B,EL}^m = 0, \quad m = 1, 2, \dots, N_B;$$

**else if** ( $TB_{EL} \leq PB_{EL}$ ) /\* perform medium-rate bit truncation \*/

20  $TB_{I\&P,EL}^n = PB_{I\&P,EL}^n, \quad n = 1, 2, \dots, N_{I\&P};$

$$TB_{B,EL}^m = PB_{B,EL} \times \frac{PB_{B,EL}^m}{\sum_{m=0}^{N_B} PB_{B,EL}^m}, \quad m = 1, 2, \dots, N_B;$$

else /\* perform high-rate bit truncation \*/

$$TB_{I\&P,EL}^n = PB_{I\&P,EL}^n + PB_{B,EL} \times \frac{PB_{I\&P,EL}^n}{\sum_{n=0}^{N_{I\&P}} PB_{I\&P,EL}^n + \sum_{m=0}^{N_B} PB_{B,EL}^m};$$

$$TB_{B,EL}^m = PB_{B,EL} \times \frac{PB_{B,EL}^m}{\sum_{n=0}^{N_{I\&P}} PB_{I\&P,EL}^n + \sum_{m=0}^{N_B} PB_{B,EL}^m};$$

5                   **endif**

**End**

[0060]     The simulation results show the effectiveness of the codecs of the present invention. Two test sequences, “Coastguard” and “Mobile,” are used in the experiments. The sequence is encoded with the (30,2) GOP structure. The BL is encoded at 384 kbps  
10     with the TM5 rate control scheme and 30 fps frame rate. The frame size is CIF 352×288. Two EL bit-planes are used in the fine prediction (i.e., the AFP and MP modes).

[0061]     Figs. 10 and 11 show the performance comparison of the method of this invention with three other methods: the baseline FGS, all-fine prediction (AFP), and the single-layer MPEG-4 codec for the two test sequences. The simulation results show that  
15     the method of this invention outperforms the other three mechanisms in a wide range of bit-rates. The AFP and the baseline FGS schemes represent two different critical bounds of quality at the highest and lowest bit-rate ranges, respectively. The purpose of the method of this invention is to find good tradeoffs between the two methods at a wide bit-rate range. This goal is achieved by adaptively introducing a predefined number of bit-

planes into the motion-compensated prediction of the BL, while slight quality degradation due to the drifting error is observed at a small range of low bit-rates (384-512 Kbps). The method of this invention is much more robust than “All-Fine” prediction.

[0062] The AFP method is applied to all B-frames which can improve the coding efficiency significantly without causing error propagation. The motion vectors are obtained using the high quality predictions. The “Inter-Layer Selection” scheme is implemented for P-frames to improve the coding efficiency at the BL and the reference frames of motion compensation may be different at both layers with the same motion information. Two sets of motion vectors for the BL and EL are not desirable because it needs much more computations and extra bit-rates for estimating and sending the extra set of motion vectors. The motion vector estimated at the BL is reused for the motion compensation at the enhancement-layer. The “All-fine” prediction suffers from about 1 dB loss when the bit rate is low. With the present invention, the quality degradation due to the drifting error at low bit-rates can be reduced significantly, while the coding gains achieved is about 1~1.5 dB than original FGS at high bit-rates.

[0063] Figs. 12 and 13 show the frame-by-frame PSNR performance comparison with a base-layer bit-rate of 384 kbps for the “Coastguard” and “ Mobile” sequences, respectively, and three different EL bit-rates: 0 kbps, 256 kbps, and 768 kbps. The scheme of this invention can reduce the drifting error more efficiently than the AFP scheme when the available bandwidth is low, while keeping the coding efficiency close to the AFP method when the available bandwidth is high. The scheme achieves significant higher PSNR quality improvement than the original FGS. Fig. 14 shows two decoded pictures using the present invention and the original FGS schemes for subjective



performance comparison.

[0064] Although the present invention has been described with reference to the preferred embodiments, it will be understood that the invention is not limited to the details described thereof. Various substitutions and modifications have been suggested in  
5 the foregoing description, and others will occur to those of ordinary skill in the art. Therefore, all such substitutions and modifications are intended to be embraced within the scope of the invention as defined in the appended claims.